PRODUCT DATA SHEET

PacBio Long Read Sequencing

Interrogate highly repetitive regions, structural rearrangements, and haplotype phasing

Circularized consensus 10kb reads, or linear continuous long reads >25kb

OVERVIEW

Long read sequencing fills a much needed gap in the genomic research toolkit, overcoming many of the shortcomings of standard short read technology, including spanning highly repetitive regions of the genome, resolving structural rearrangements, and providing robust haplotype phasing information. Long read sequencing allows for better insight into oncogenics, vaccine development, de novo assembly and characterization of genetic disease.

Broad Genomic Services provides long read sequencing capabilities in the form of PacBio's Sequel II platform,and is available in a couple of configurations; Circular Consensus Sequencing (CCS), and Continuous Long Read (CLR) sequencing.

Circular Consensus Sequencing

Utilizing 10 kb circularized library molecules and resequencing reads typically ⩾60 kb in length, we can effectively provide ~6x unique read coverage. This allows us to use CCS error correction to achieve high fidelity reads with error rates of 1% or less.

Continuous Long Reads

Alternatively, we can provide linear single molecule long read data using the CLR library preparation and sequencing option, by size selecting library fragments ⩾30 kb, resulting in exceptionally long reads, with some reaching over 150 kb. Overall yields on the Sequel II platform represent an ~ 8x increase in the amount of data per SMRT cell compared to previous Sequel I instrument, and is comparable to ~7-8x coverage of whole human genome.

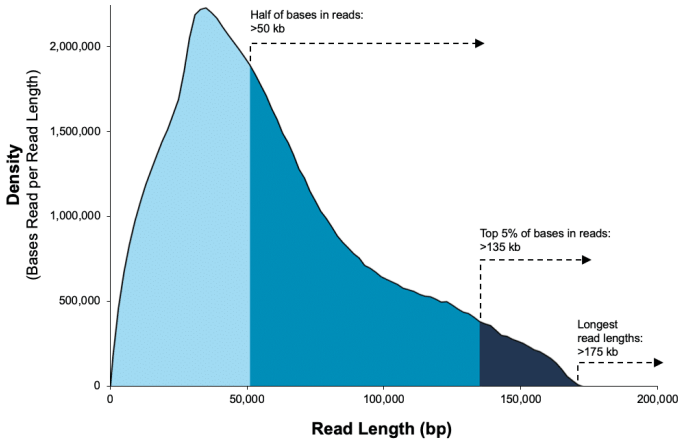
WHAT'S INCLUDED

- Sample Receipt, incoming QC
- Library Construction (10 kb CCS or >30 kb CLR)
- PacBio Sequel II Sequencing
- Data Analysis and Delivery

FOR MORE INFORMATION

Web: genomics.broadinstitute.org
Email: genomics@broadinstitute.org

BASES PER READ BY READ LENGTH



Bases per Read by Read Length resulting from CLR library construction and sequencing. Half of all base reads come from bases over 50kb in length. 5% of bases are found in reads over 135 kb.

PERFORMANCE SPECIFICATIONS

	CCS	CLR
Mean Insert Size:	9.5 kb	25 kb
Error Rate:	0.60%	10-15%
Q-Score:	Q22	Q10-15
WGS Coverage (per cell):	8x	12x

INPUT REQUIREMENTS

- >5 ug high quality genomic DNA at 16ng/uL minimum concentration
- Minimum Sample data including collaborator participant ID, collaborator sample ID, gender

DATA DELIVERABLE

- Data accessed via secure online digital transfer
- CCS error correction for all circularized samples
- Aggregated hg38 BAM (hg19 on request)
- 4 Variant Call Files (VCFs)
  - HaplotypeCaller
  - DeepVariant
  - PacBio Structural Variant (PBSV) - long read SV caller
  - Sniffles - long read SV caller

BROAD INSTITUTE

GENOMIC SERVICES